Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L | week 12

Pierre-Luc Germain

ETH Zürich

Plan for today

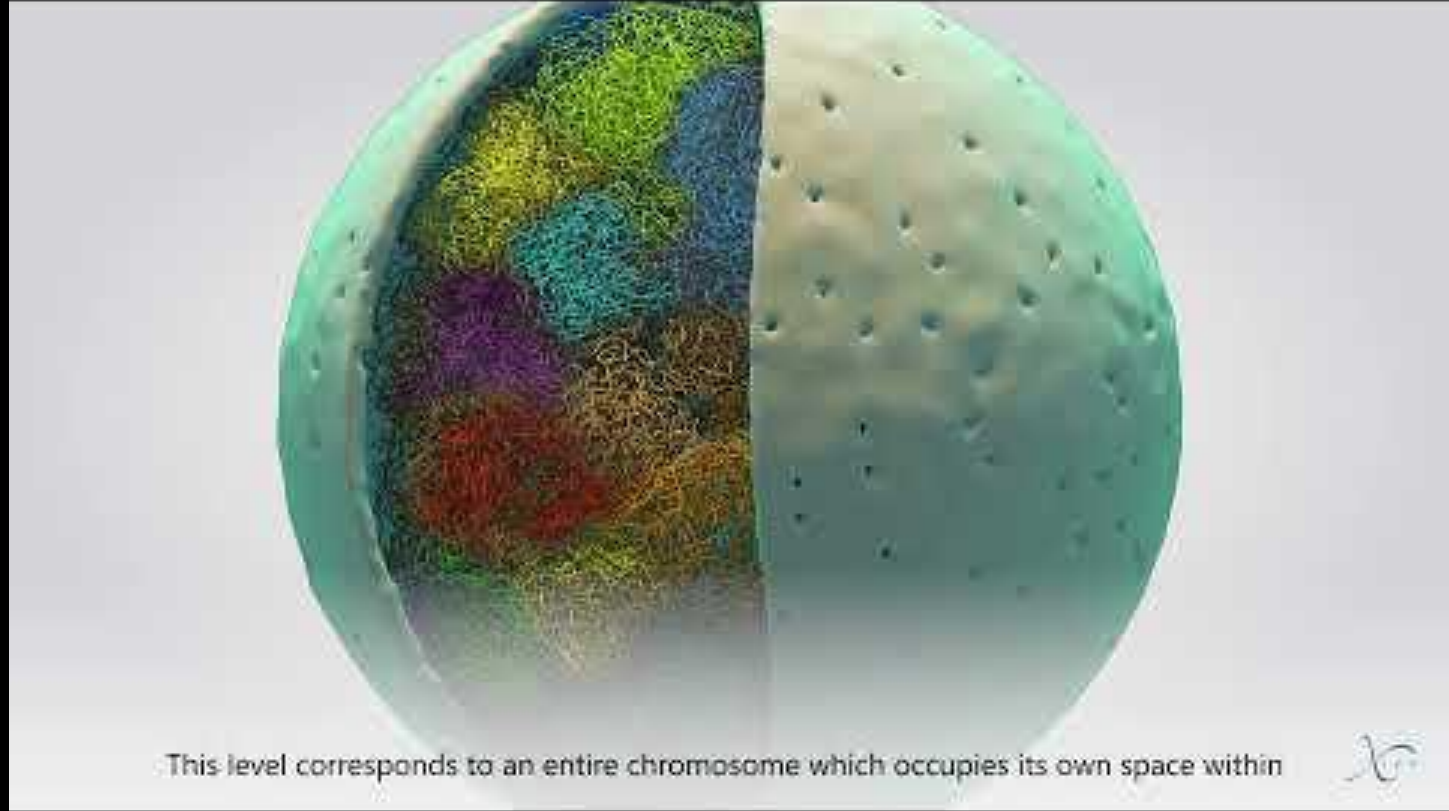- Last week's assignment -> pushed to next week

- **Theory:**

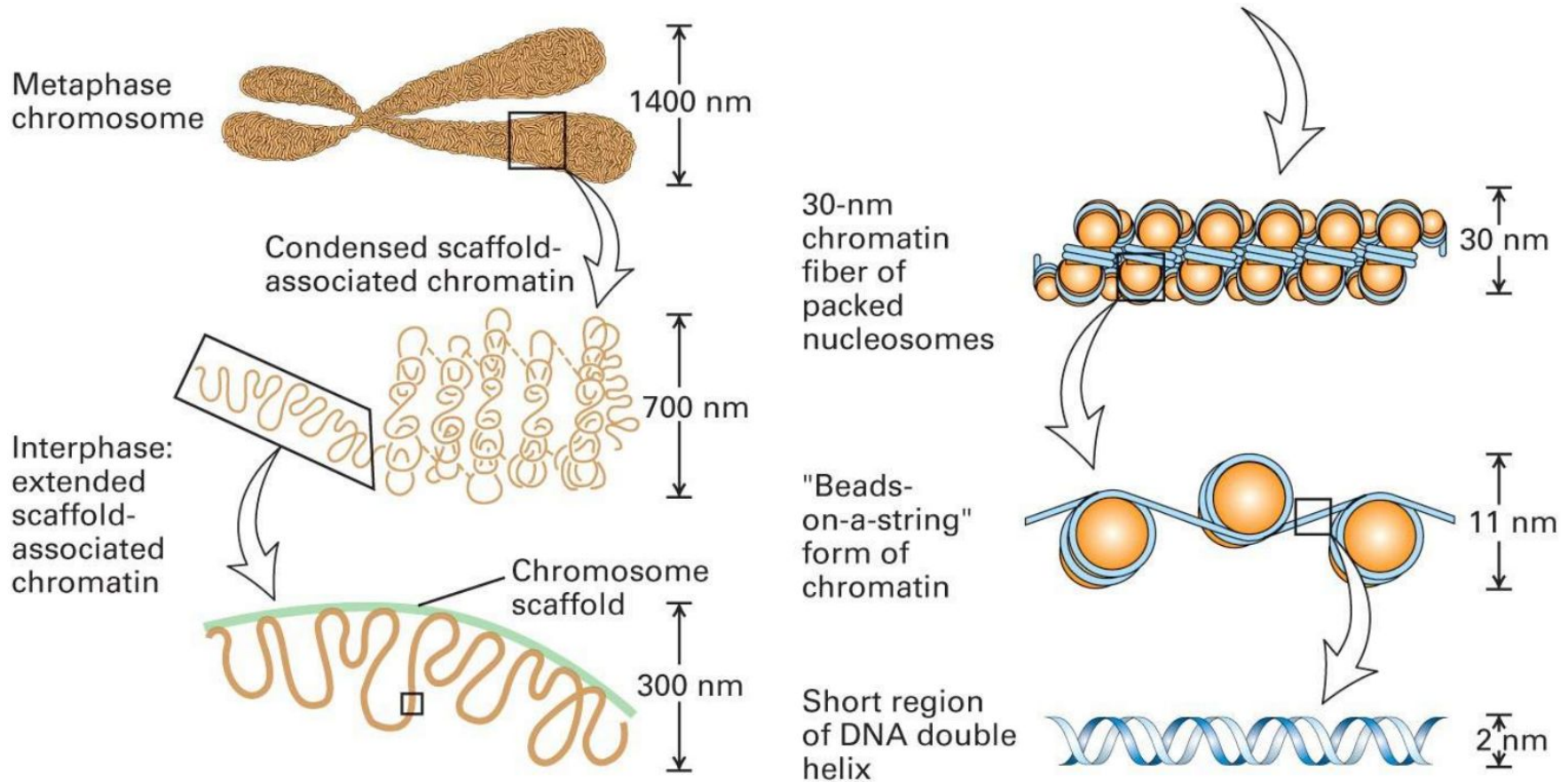 Chromatin conformation & related technologies

- **Practice:**

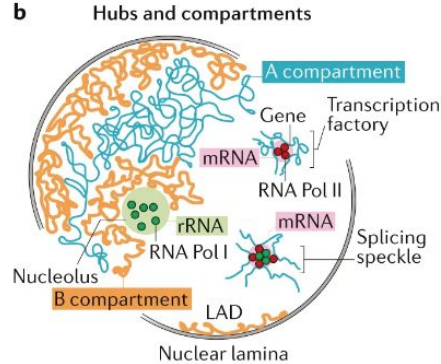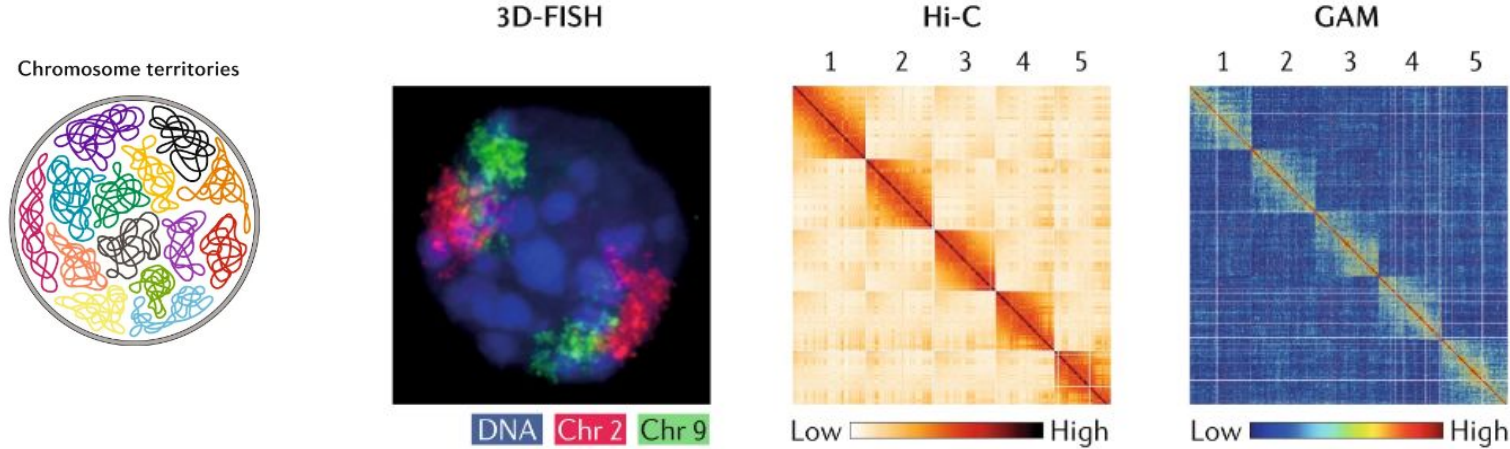 Using long-range interactions to annotate distal sites

3D organization of the genome

This level corresponds to an entire chromosome which occupies its own space within

Metaphase chromosome — 1400 nm

Condensed scaffold-associated chromatin

Interphase: extended scaffold-associated chromatin

Chromosome scaffold — 700 nm

300 nm

30-nm chromatin fiber of packed nucleosomes — 30 nm

"Beads-on-a-string" form of chromatin — 11 nm

Short region of DNA double helix — 2 nm

Lodish et al. *Molecular Biology of the Cell* (5th ed.). W.H. Freeman & Co., 2003.

The nucleus is organized into chromosome territories

Chromosome territories

3D-FISH

DNA Chr 2 Chr 9

Hi-C

1 2 3 4 5

Low ▭ High

GAM

1 2 3 4 5

Low ▭ High

b Hubs and compartments

A compartment
Gene
Transcription factory
mRNA
RNA Pol II
rRNA
mRNA
RNA Pol I
Splicing speckle
Nucleolus
B compartment
LAD
Nuclear lamina
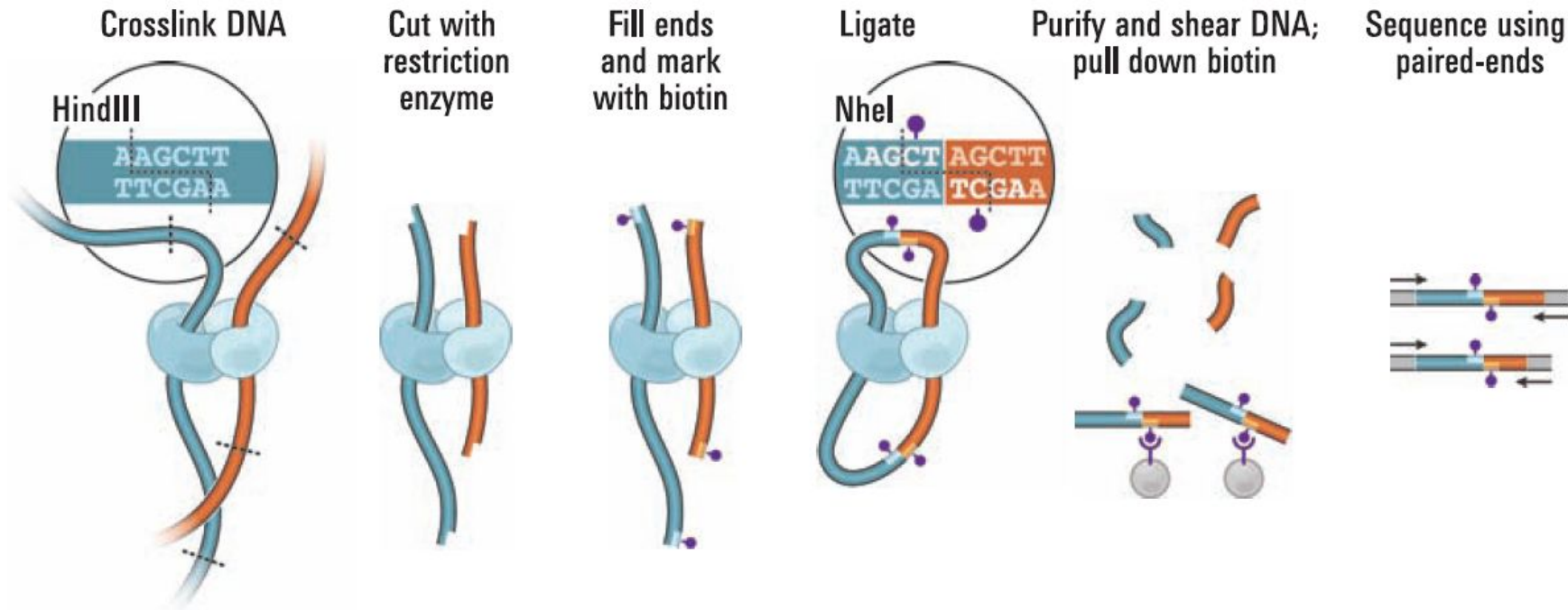
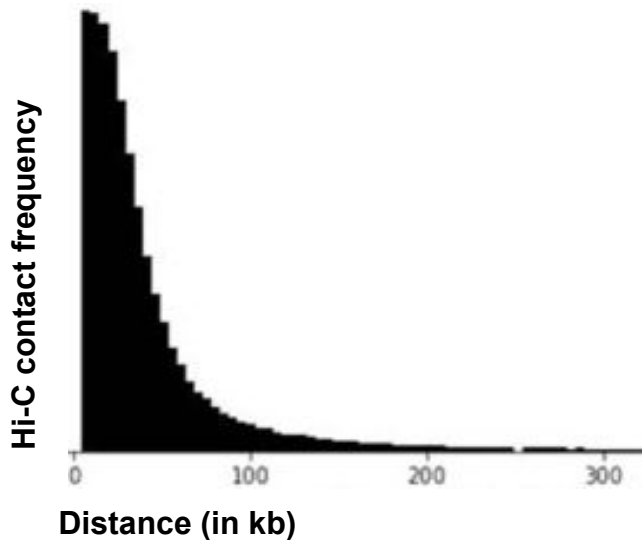Electron spectroscopy imaging
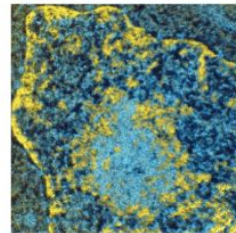
Heterochromatin
Euchromatin

Chromosomes are split into active (A) and inactive (B) compartments

(Adapted from Kempfer and Pombo 2020)

# Sequencing-based conformation capture

(Lieberman-Aiden et al. 2009)

Sequencing-based conformation capture

Hi-C contact frequency vs. Distance (in kb)

Under the null hypothesis of random 3D organization, we expect to have more contacts from DNA regions that are closer to each other.
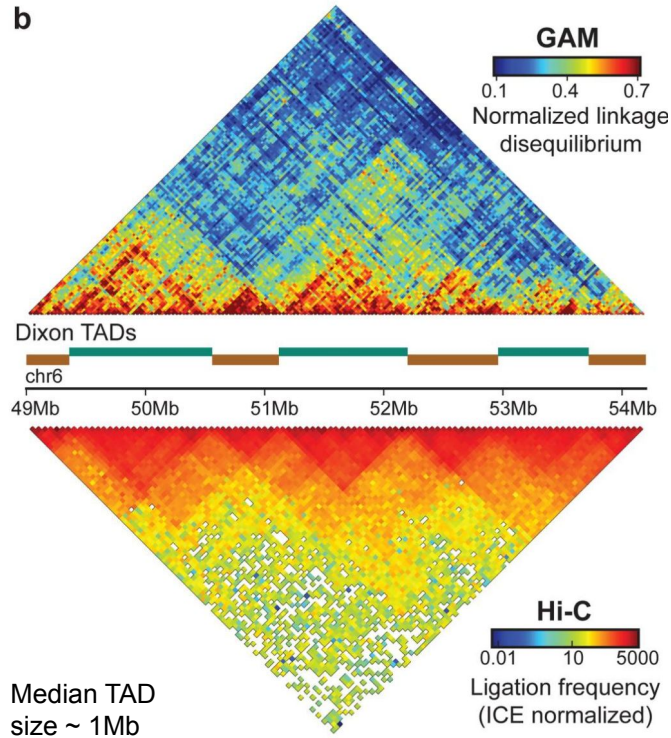
This is why, when identifying interactions, methods look for an excess over this null hypothesis
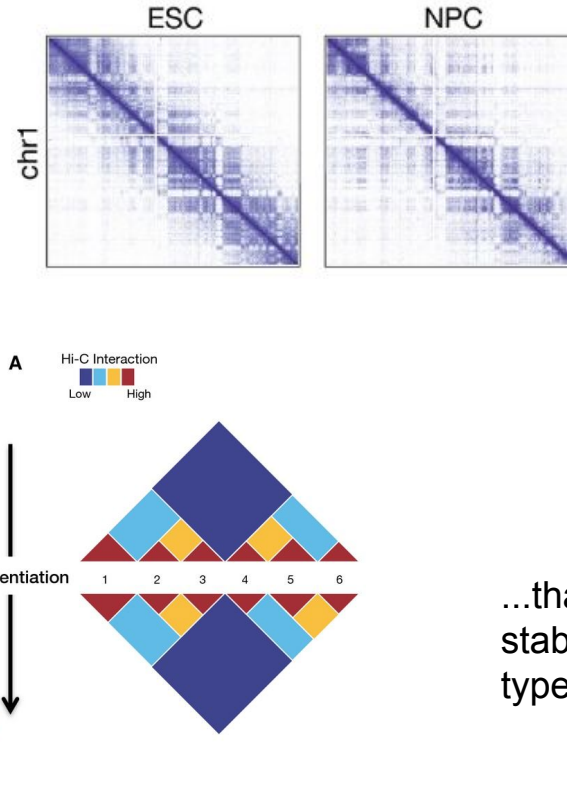
Sequencing-based conformation capture
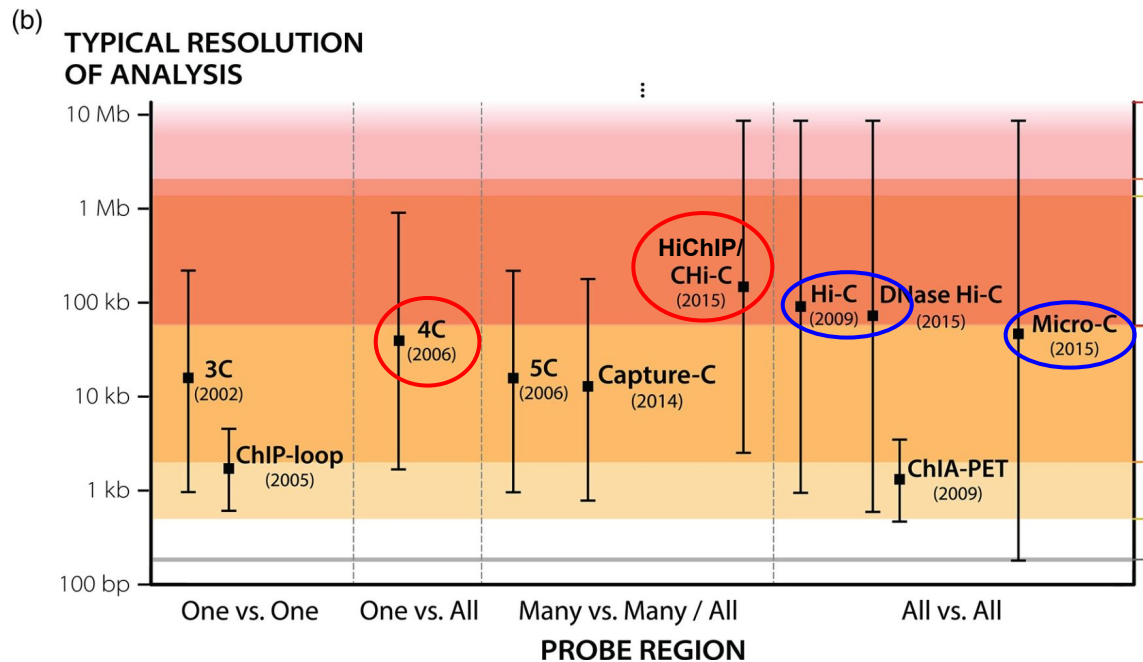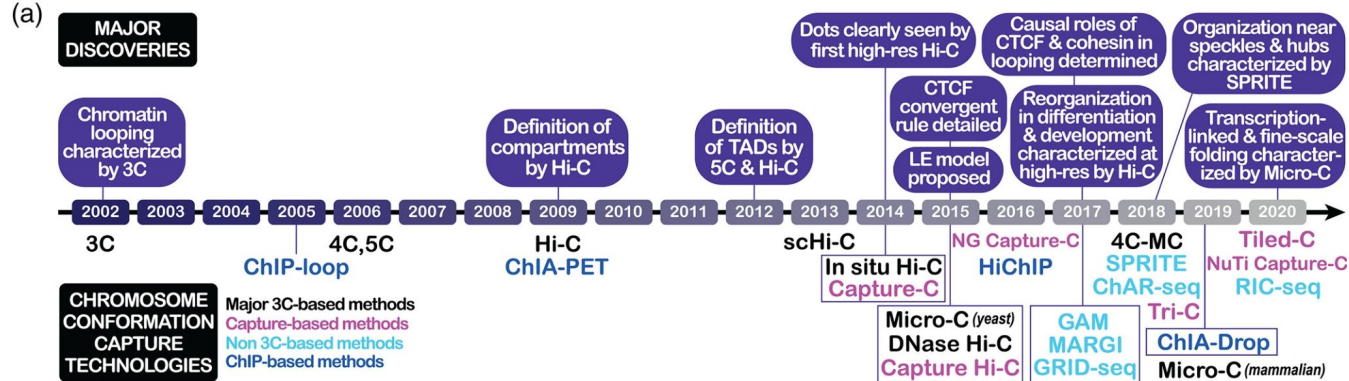
(Adapted from Goel and Hansen 2020)

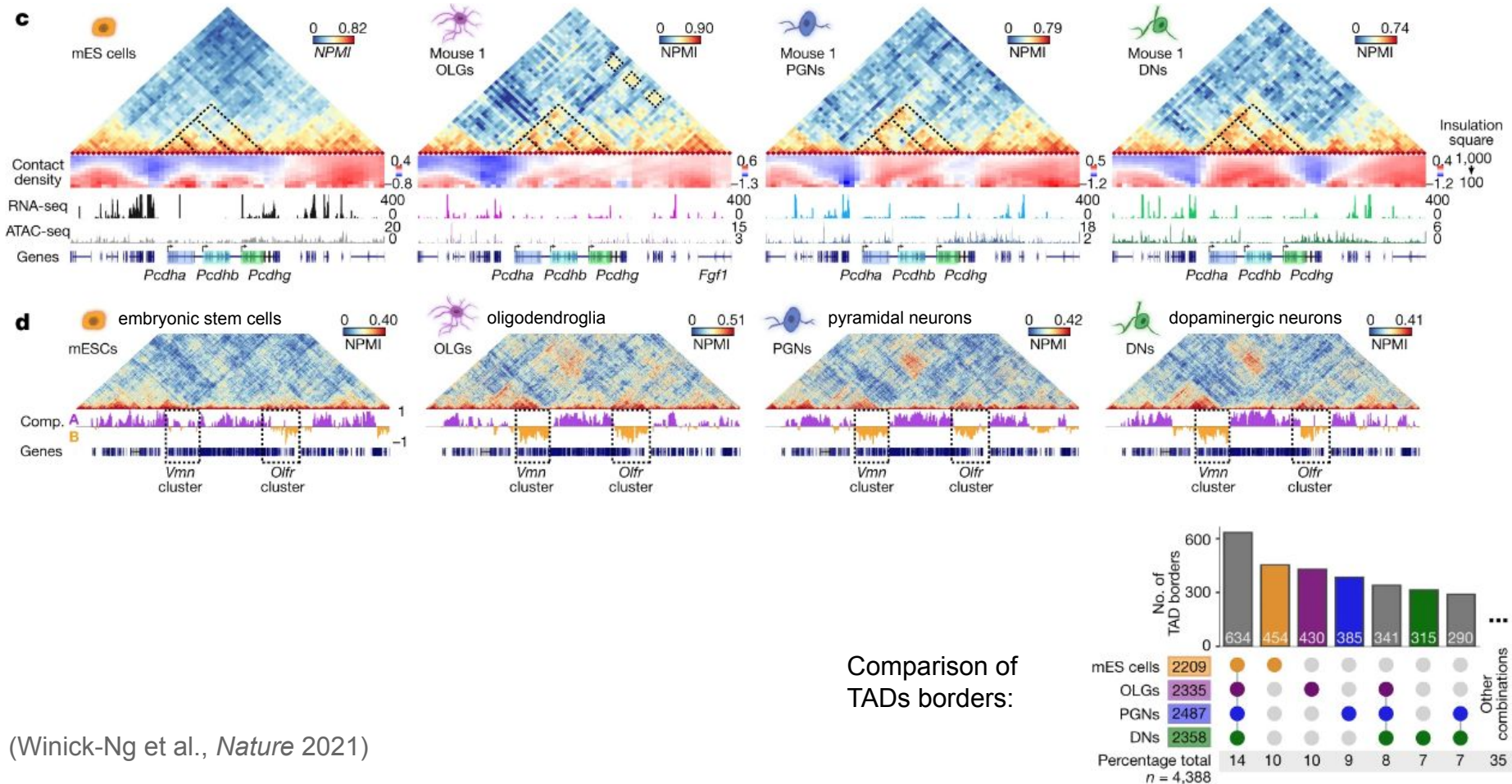# Chromosomes are organized into topologically associated domains (TADs)
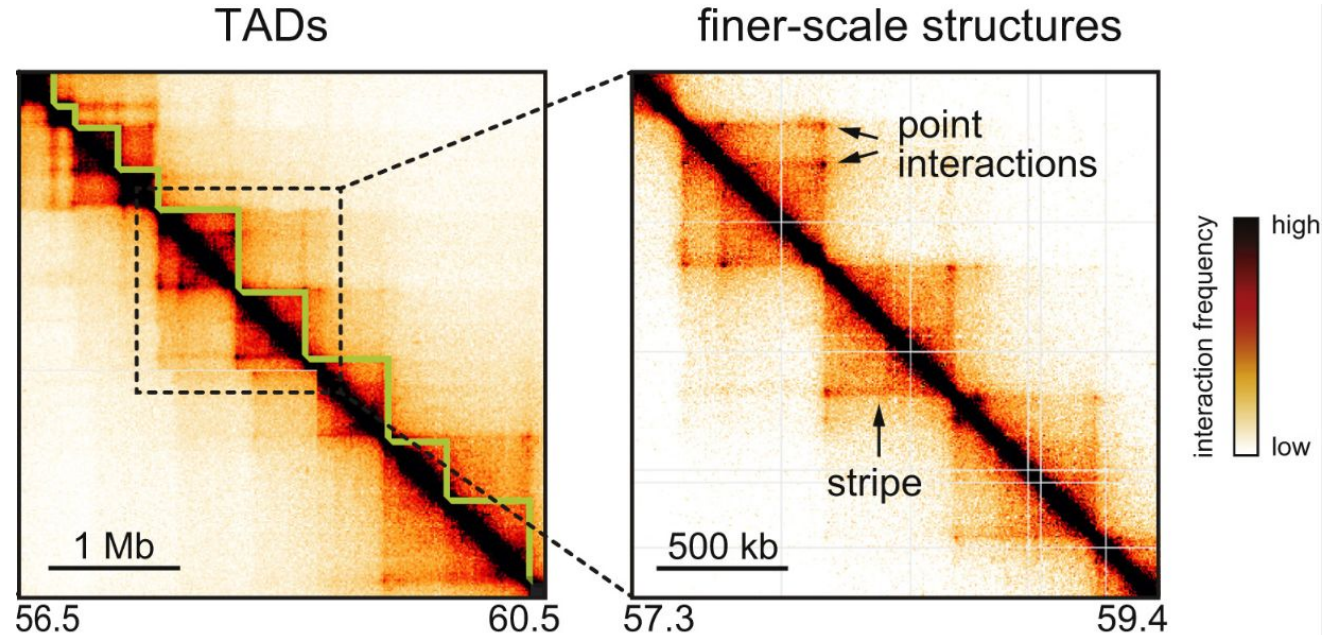

(Beagrie et al., 2017)

(Fraser et al., 2015)

...that are rather stable across cell types

TADs rearrangement across cell types

c

mES cells — NPMI 0 0.82
Mouse 1 OLGs — NPMI 0 0.90
Mouse 1 PGNs — NPMI 0 0.79
Mouse 1 DNs — NPMI 0 0.74

Contact density
RNA-seq
ATAC-seq
Genes

Pcdha Pcdhb Pcdhg
Pcdha Pcdhb Pcdhg Fgf1
Pcdha Pcdhb Pcdhg
Pcdha Pcdhb Pcdhg

Insulation square 1,000 100

d

embryonic stem cells — mESCs — NPMI 0 0.40
oligodendroglia — OLGs — NPMI 0 0.51
pyramidal neurons — PGNs — NPMI 0 0.42
dopaminergic neurons — DNs — NPMI 0 0.41

Comp. A B
Genes

Vmn cluster *Olfr* cluster

Comparison of TADs borders:

							Other combinations
No. of TAD borders	634	454	430	385	341	315	290	...
mES cells 2209	●	●	●	●	●	●	●	
OLGs 2335	●	●	●	●	●	●	●	
PGNs 2487	●	●	●	●	●	●	●	
DNs 2358	●	●	●	●	●	●	●	
Percentage total *n* = 4,388	14	10	10	9	8	7	7	35

(Winick-Ng et al., *Nature* 2021)

TAD boundaries are defined by very stable point interactions…

…which represent CTCF binding sites

TADs

finer-scale structures

point interactions

stripe

interaction frequency

high

low

1 Mb

500 kb

56.5   60.5   57.3   59.4

(Adapted from McCord, Kaplan and Giorgetti, Mol Cell 2020)

CTCF forms **convergent** dimers at loops

CRISPR-INVERSION

CTCF Forward
CTCF Reverse
Cohesin
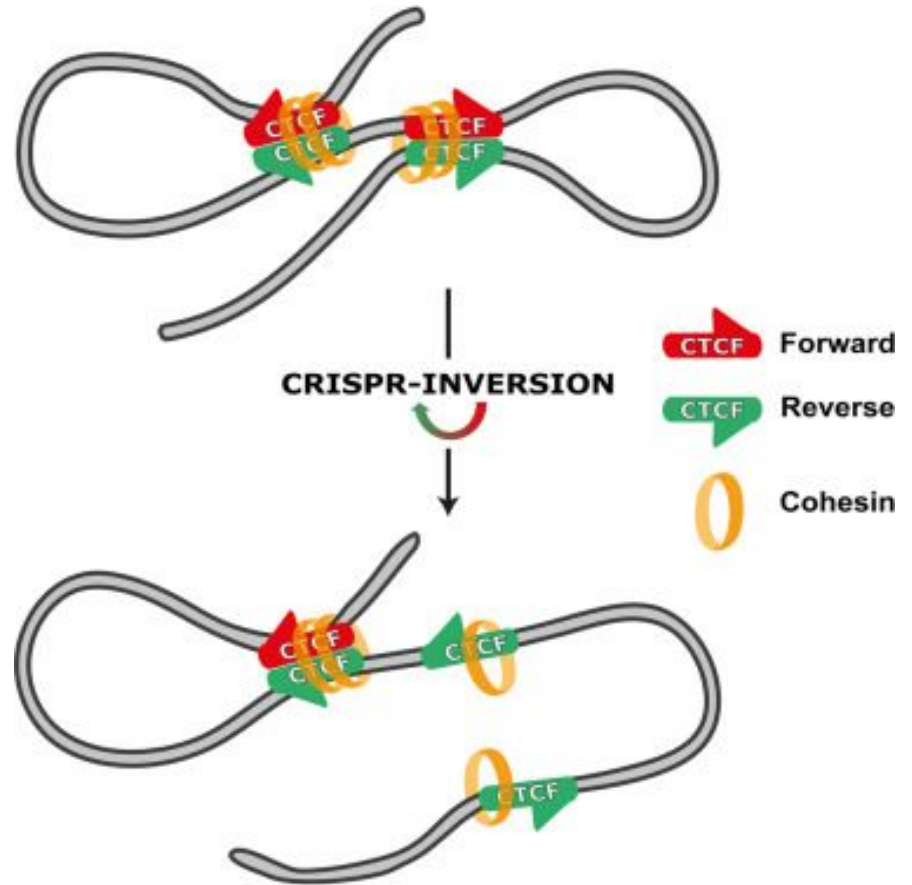
(Adapted from de Wit et al., Mol Cell 2015)

The loop extrusion model

CTCF dimers form loops
by blocking Cohesin
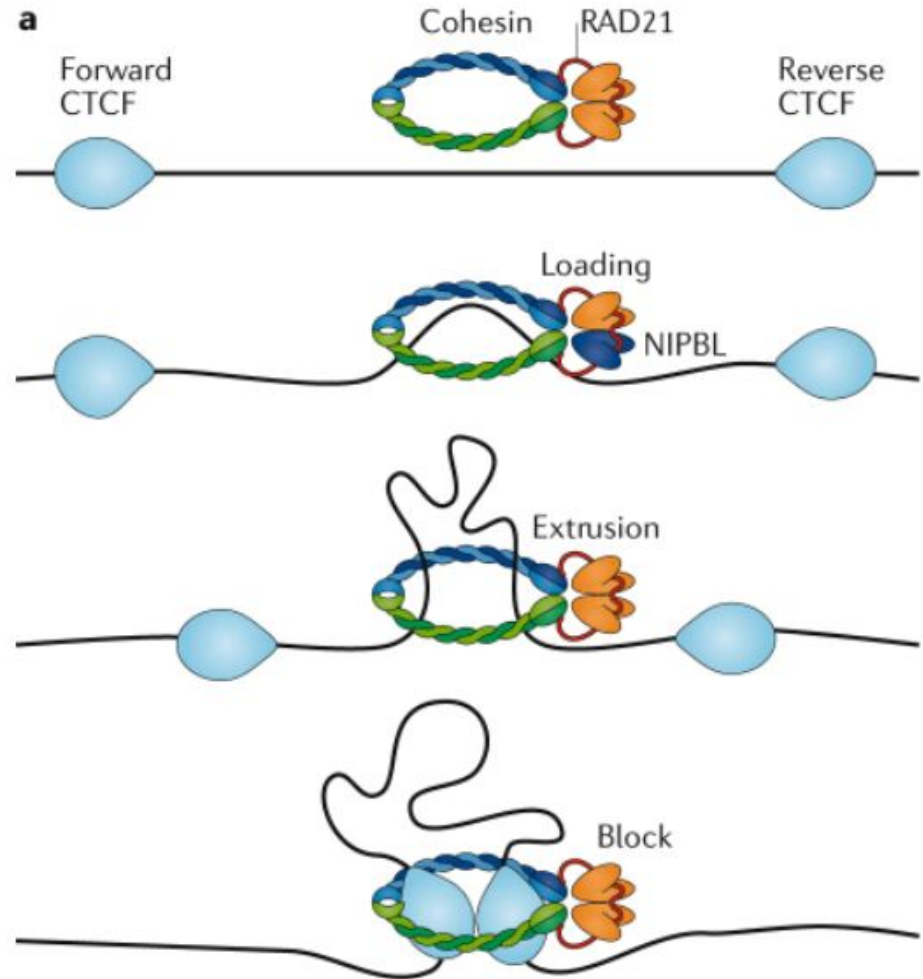
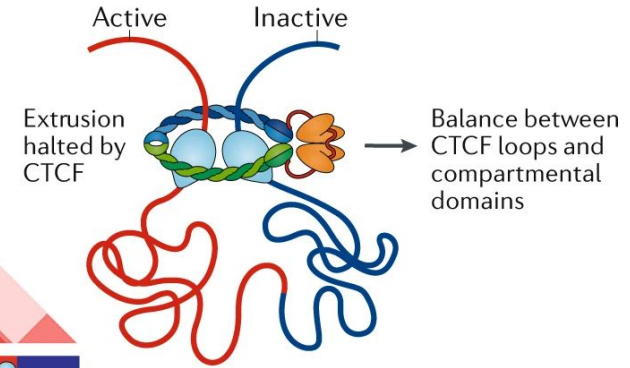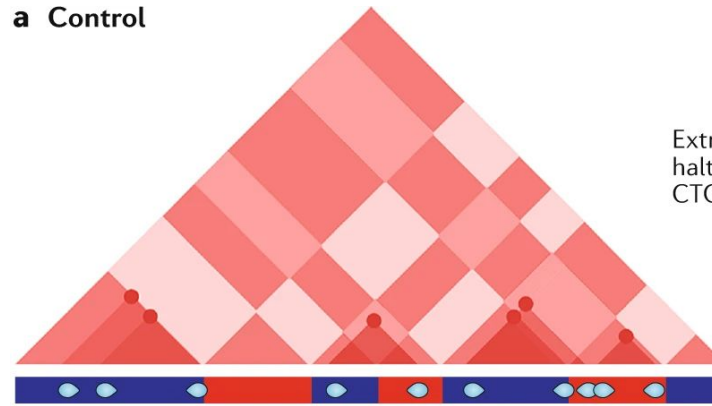CTCFs dynamically exchange at their binding sites on DNA.

CTCF depletion disrupts especially TAD-internal structures

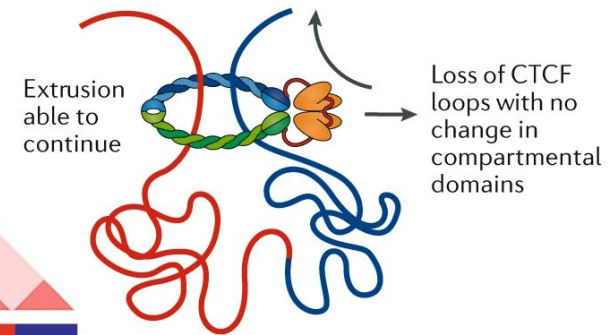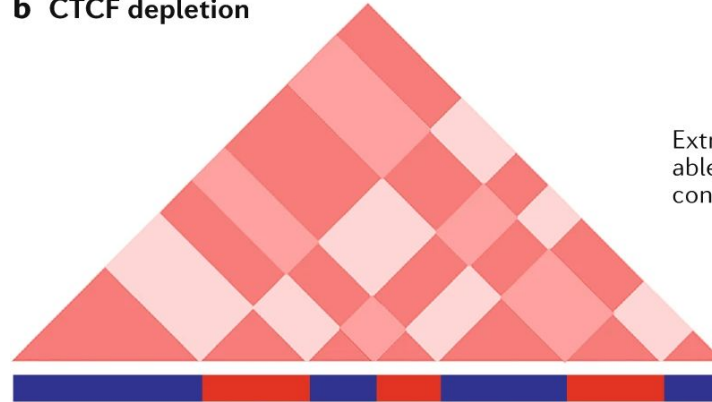presumed to include promoter-enhancer loops
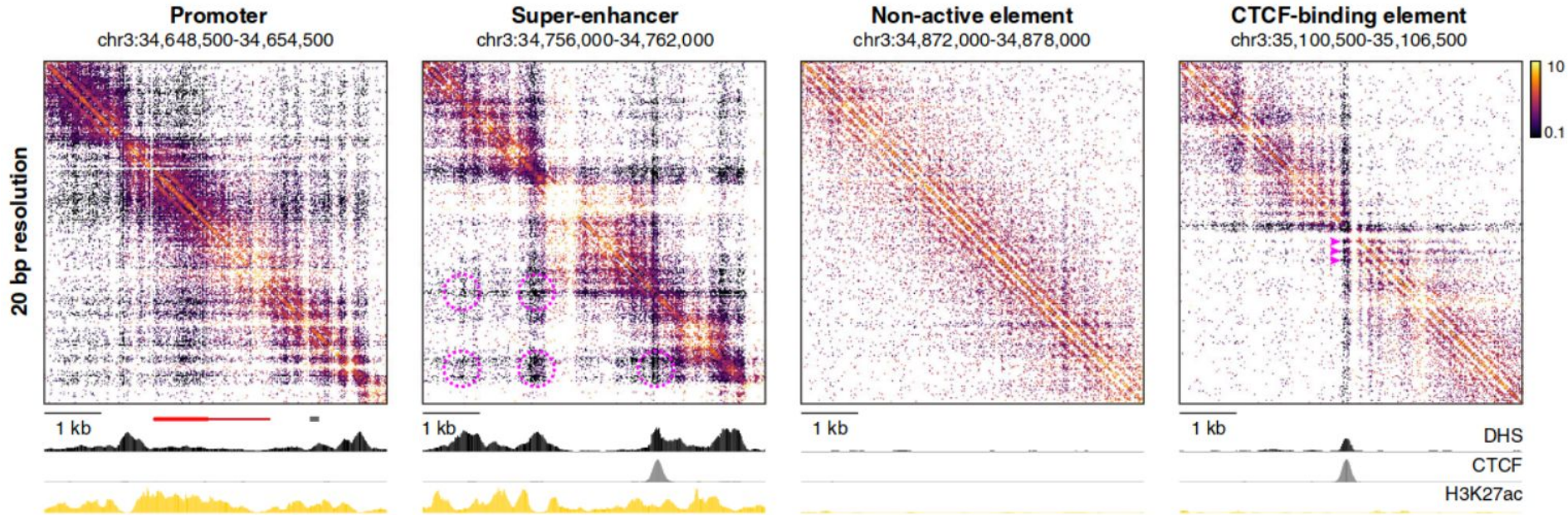
however…

a Control

b CTCF depletion

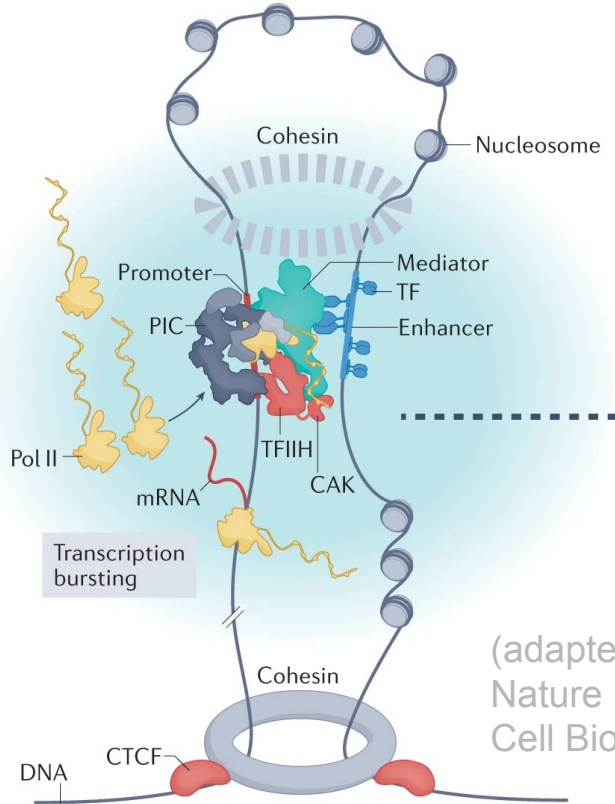Active · Inactive · ⬭⬭ CTCF loop anchors

When going at very high resolutions, similar contact patterns appear, that are however not associated with CTCF

The structures however appear correlated with DHS, suggesting that at this fine-grained level the it is determined by (most likely combinations of) other factors

(20bp-resolution contact maps, adapted from Aljahani et al., *Nat Comm* 2022)

According to the prevailing model, enhancers physically interact with target TSS through combinations of proteins (in particular *Mediator*)

An alternative view rests on diffusion between regulatory elements

Cohesin — Nucleosome
Promoter — Mediator
— TF
PIC — Enhancer
Pol II
TFIIH
mRNA — CAK
Transcription bursting
Cohesin
DNA — CTCF

(adapted from Richter et al., Nature Reviews Molecular Cell Biology 2022)

In both views, chromatin loops create local neighborhoods in which these interactions can take place

D
Pol II complex
Weak E
Strong E
H3K27ac
TF/Cofactors
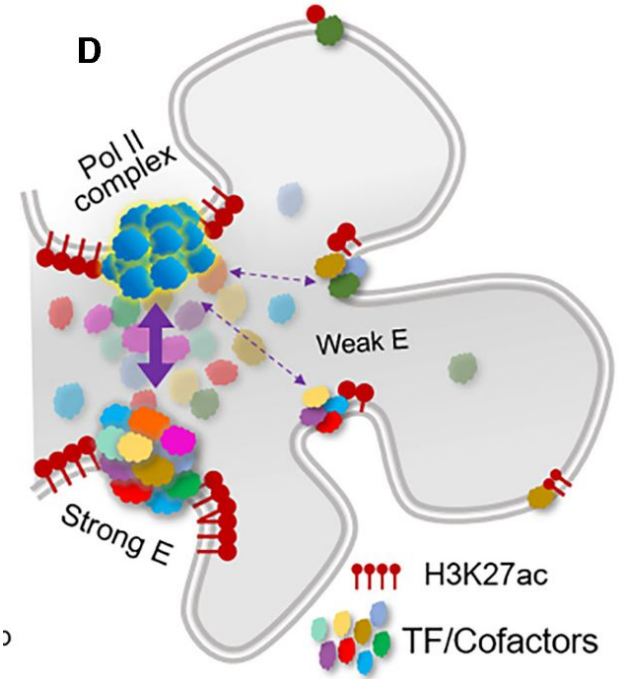
(adapted from Zhu et al., Nucleic Acids Research 2021; see also Galouzis and Furlong, Current Opinion in Cell Biology 2022)

Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1

Tsung-Han S. Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S. Hansen, Xavier Darzacq ✉ & Robert Tjian ✉
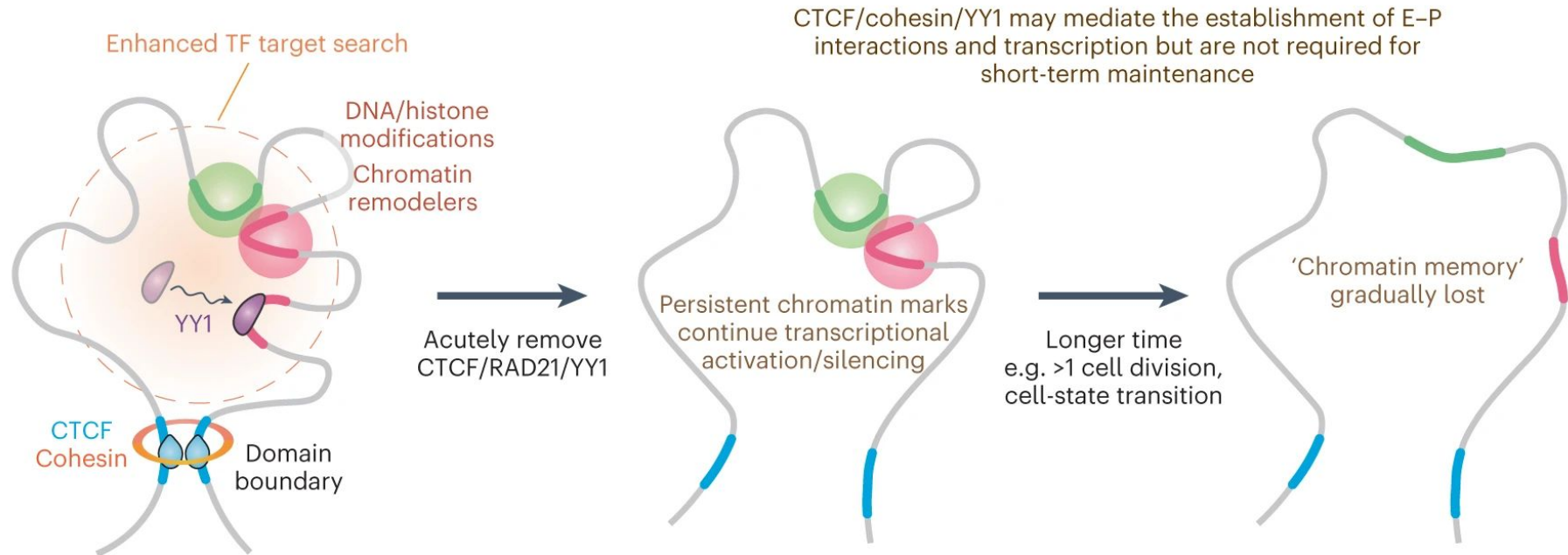
Proximity is not sufficient for enhancer-mediated activation

Article

## Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome

Miguel Martinez-Ara [1], Federico Comoglio [1,2], Joris van Arensbergen [1,3], Bas van Steensel [1,4]
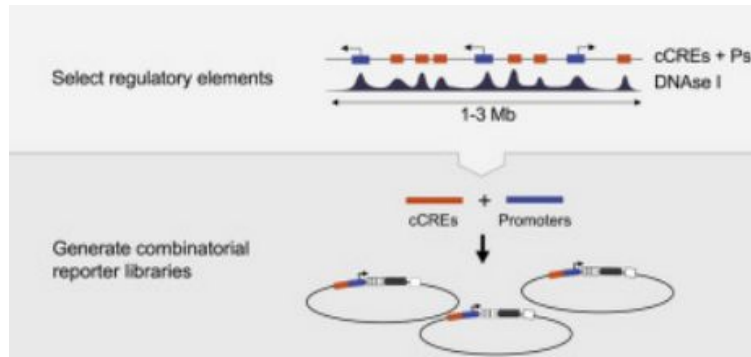
- Even when enhancers are made to come in proximity to a promoter, **not all enhancers can regulate all promoters**

- E-P compatibility (i.e. the specificity of the interactions) appears to be dictated by TF interactions between the two domains
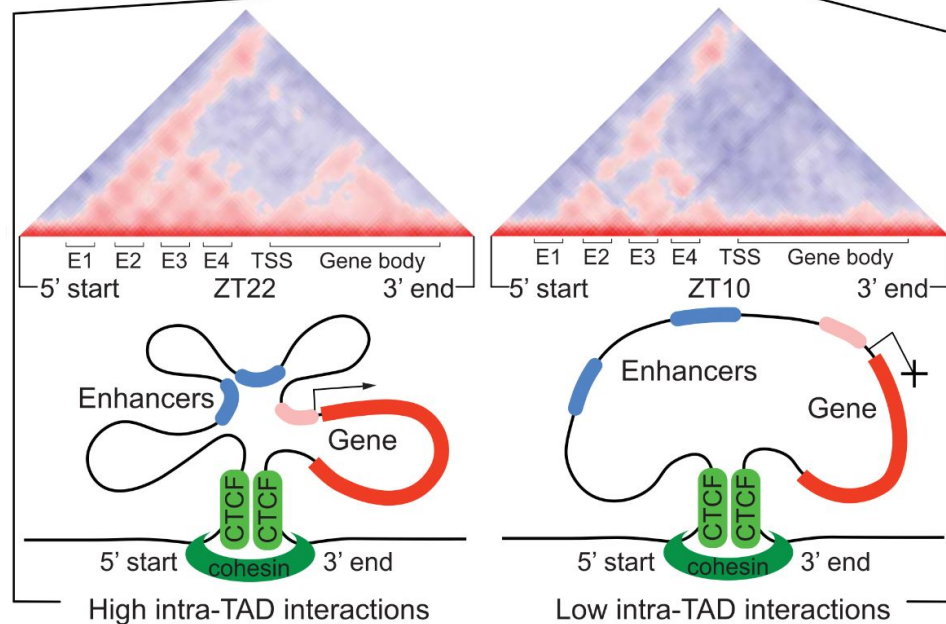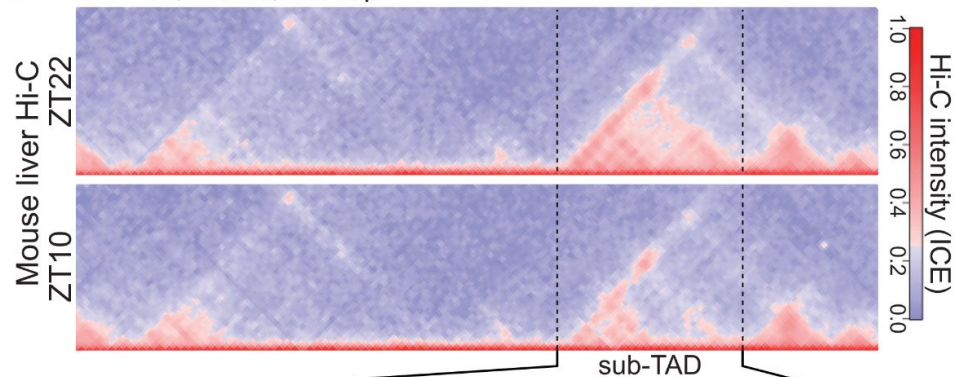
Changes in chromatin contact occur within stable TADs/subTADs

Here an example driven by the circadian rhythm (mouse liver, ZT22=5am ZT10=5pm)

(Kim et al., Science 2018)

A few take-home messages

- the nucleus is organized into **chromosome territories**,
  and chromosomes are organized into active/inactive **compartments**

- chromosomes are further organized into **TADs** that are
  - largely stable across cellular states and
  - are delimited by CTCF dimers (loops)

- chromatin loops bring regions of chromatin in proximity,
  - enabling the formation of enhancer-promoter interactions,
  - but are not required for the short-term *maintenance* of these interactions

- Functional enhancer-promoter interactions appear determined by **TF interactions**, made possible by loops

Sequencing-based conformation capture
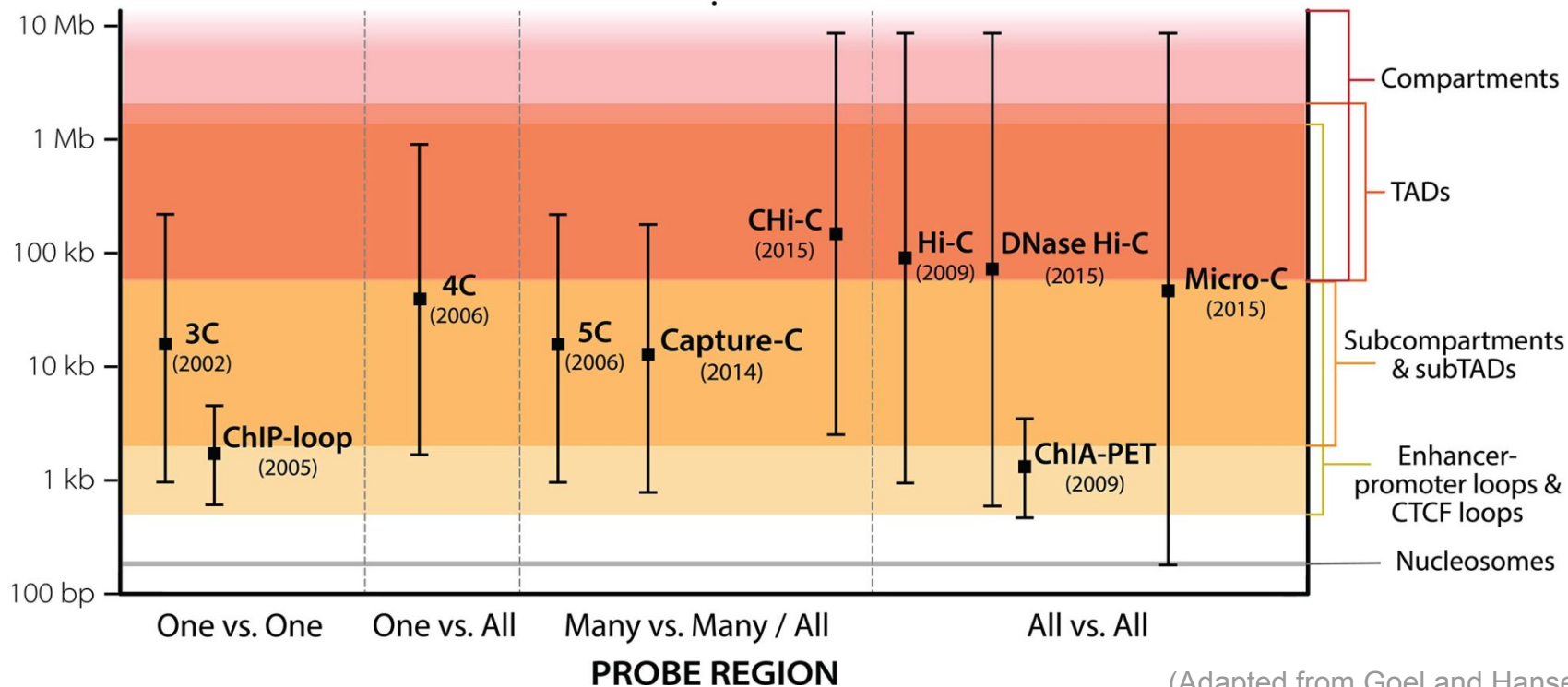
(Adapted from Goel and Hansen 2020)

(See also Hun Lee, Wu and Rhie, Epigenetics & Chromatin 2022 for a review)

Data analysis - some references

**Method &
analysis refs**

**Main output &
formats**

Visualization in R

HiCBricks

plotgardener

- Hi-C:
  - Juicer

- Contact maps
- TADs/
compartments:
  - bedpe (?)

Data structure in R

interactionSet

- HiChIP:
  - Dovetail doc
  - FitHiChIP

- Interactions:
  - bedpe / pairix
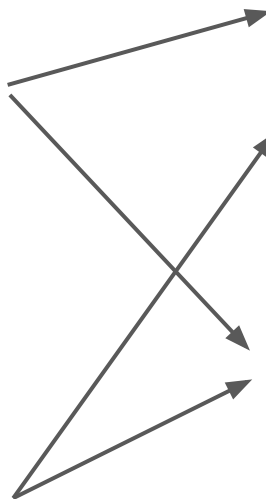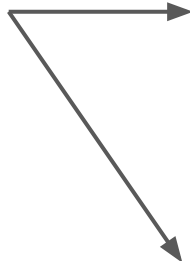
Practical

Enhancer-promoter interactions:
what to do with distal regulatory elements?

- Enhancers within <1kb of a TSS nearly always regulate that TSS (Sanyal et al., Nature 2012; Morris et al., Science 2023),

- Beyond 2.5kb, more than half of the interactions are not with the closest gene (Sanyal et al., Nature 2012).

- In order to link such distal regulatory regulatory elements to target genes, we have two options:

  - Use interactions obtained from (cell-type specific) chromatin conformation data

  - Use generic enhancer-target predictions (e.g. Salviato et al., NAR 2021)

Assignment

- Choose a transcription factor (e.g. p300), and obtain peaks from ENCODE (ChIP-seq in a human context!)

- Isolate the peaks that are:

    - Between 2.5kb and 10kb from a TSS

    - More than 10kb from a TSS

- For each set of peaks:

    - Subset to those peaks that have a predicted distal target(s) using Salviato et al. (2021)

        - You can download a GRanges of those interactions at

            https://ethz-ins.org/content/hg38.SalviatoDistalEnhancerTargets.GR.rds

    - Find the nearest TSS for each peak

    - **In what proportion of the cases is the predicted target the closest gene?**

- Hints:

    - you can use the `annotateRegions` function, as we did in week 4, to get the gene nearest to each peak

    - beware not to count, when calculating proportions, peaks that don't have interactions with any TSS!

- Expected for of the answer:

    - "Of the genes that are between 2.5 and 10kb from the nearest TSS, XX % form an interaction with that nearest gene. Of the genes that are more than 10kb away from the nearest TSS, XX % form an interaction with that nearest gene."